

MemoQ – új megközelítés a fordítástámogatásban

Lengyel István, Kis Balázs, Ugray Gábor

Kilgray Kft.

{istvan.lengyel;balazs.kis;gabor.ugray}@kilgray.com

A fordítók körében a fordítástámogatási eszközök használatáról végzett felmérések (Drugan 2004, Fulford-Granell-Zafra 2004, Somers 2003) kimutatták, hogy csak kevés fordító alkalmaz teljes egészében számítógéppel támogatott munkafolyamatot. Ha azonban a fordítókat olyan közösségnek tekintjük, amelynek tagjai hasonló módon és hasonló szabályok, feltételezések alapján dolgoznak, és a fordítási folyamatot ennek alapján közelítjük meg, a termelékenységet tovább növelhetjük. A gyakorlatba ez holisztikus szemlélettel megvalósított, integrált fordítástámogató szoftver segítségével ültethető át. A MemoQ projekt az előadásban ismertetett megközelítés első implementációja.

1. Bevezetés

A legtöbb tudományos cikk a fordítást önállóan végzett tevékenységnek tekinti, és mint ilyent, a külvilágtól elszigetelve vizsgálja. A valóságban azonban a fordítók a piac törvényei szerint dolgoznak, ahol a hatékonyság még a minőségénél is fontosabb. A hatékonyság a hasznosság és a költségráfordítás közötti kapcsolatot határozza meg. A hasznosság a fordítás felhasználásából származik. Így a legjobb fordítás – a piaci viszonyok alapján – az a fordítás, amely teljesen megfelel egy előre meghatározott, kimondott vagy ki nem mondott célnak.

A számítógéppel támogatott fordítás célja a hatékonyság növelése. Mindazonáltal a legtöbb ilyen rendszert lépcsőzetesen építették fel, és a fejlesztés első szakaszában mindegyik csupán az egyéni, íróasztala mögött egyedül dolgozó fordító igényeit elégítette ki. Az Internet később elsődleges kommunikációs csatornává és a leggyorsabb kutatási forrássá vált a fordítók számára, és mindez jelentősen csökkentette a piaci reagálási időt.

Az Internet azonban nem csupán a fordítási munkát alakította át, hanem további fordítási igényeket is támasztott. A fordítók kiszakadtak kreatív magányukból, és olyan helyzetbe kerültek, ahol minden és mindenki kapcsolatban van egymással, a határidők pedig sokkal szorosabbak, mint korábban.

A számítógépes fordítástámogató eszközök (*Computer-assisted Translation, CAT*) fejlesztői természetesen reagáltak az internet kihívására, de eszközeik – mivel nem írták át teljesen őket – nem tükrözik a paradigmaváltást. A fordítómemóriák továbbra is egygépes alkalmazások, minden felhasználónak egyforma lehetőségek állnak rendelkezésére.

A programok nem veszik figyelembe, hogy a fordítási projekten nemcsak fordítók dolgoznak, hanem projektmenedzserek, nyelvi vezetők, lektorok, kiadványszerkesztők

is. Minden felhasználónak mélységében kell ismernie a fordítómemóriákat, ha hatékonyan akar dolgozni, és a fordítómemóriák (*translation memory*, TM) elérhetőségében sem állt be nagy változás.

Amellett érvelünk, hogy hálózati környezetben lehetséges a munkamegosztás előnyeinek kihasználása, egyszerűsíthető a felhasználói felület, megakadályozható az adatvesztés – azáltal, hogy kevésbé kötjük a fordítómemóriát az operációs rendszer fájlrendszeréhez. Mindez pedig jobb munkakörülményeket és végső soron megtakarítást eredményez.

A 2. részben leírjuk, milyen előnyökkel jár, ha a fordítási munkafolyamatot vesszük kiindulási alapul a fordítástámogató eszköz fejlesztésekor. A 3. rész a fájl- és könyvtárkezelés kiiktatásáról szól. A fájl- és könyvtárkezelés véleményünk szerint az a terület, amely a leginkább csökkentheti a hatékonyságot, a meglévő erőforrások fel nem használása révén. A 4. részben a rendszer fejlesztésének lehetőségeit tárgyaljuk. Az 5. rész összegzi a MemoQ alapelveit.

2. A fordítási munkafolyamat kezelése

A fordítás projektszemléletű tevékenység. Az amerikai Projektmenedzsment Intézet (*Project Management Institute*) 2002-es definíciója szerint „a projekt ideiglenes erőfeszítés egy adott cél elérése érdekében. A projektek abban különböznek az üzletviteltől, hogy az üzletvitellel kapcsolatos tevékenységek folytonosak és ismétlődnek, míg a projektek időszakosak és egyediak.” A projektnek tehát mindig van kezdete és vége – akkor is, ha egyszerűen csak megszakítják. A fordítási projektek esetében egy vagy több szereplő működik együtt annak érdekében, hogy a forrásdokumentum(oka)t olyan célnyelvi dokumentummá alakítsa, amely legalább egy nyelvi és kulturális kontextusban megfelelő. A megbízás részletes elemzése után a fordítási projektek különböző tevékenységekre oszthatók, amelyeket más és más személy végezhet.

Robichaud és L'Homme (2003) egyszerűsített listát közöl a munkafolyamat lépéseiről, amely felhasználható a fordítástámogatási és gépi fordítási eszközök használatának tanításához is:

1. A forrásnyelvi szöveg átvétele
2. Első olvasat
3. Terminológiai- és háttérkutatás
4. Fordítás
5. Lektorálás és javítás
6. Átnézés, egységesítés, olvasószerkesztés
7. A célnyelvi szöveg átadása

A munkafolyamat kis számú lépésből épül fel, de a lépések sorrendje változhat, a lépések pedig ismétlődhetnek, így rendkívül sokféle munkafolyamat képzelhető el, főleg a hosszú szövegek fordítása és a szoftverlokalizálás során. Ennek ellenére már a munkafolyamat legelején tudjuk, hogy

- (1) kik dolgoznak a projekten,
- (2) milyen erőforrásokat használunk fel a projekt során.

A fordítóirodák a legtöbb esetben projektmenedzsert alkalmaznak, akinek a tevékenysége átfogja a teljes munkafolyamatot. A projektmenedzser felelős a minőségbiztosításért, ő biztosítja a munkafolyamat szereplői (fordítók, lektorok, szakértők, kiadványszerkesztők stb.) közötti problémamentes információáramlást, legfőbb feladata pedig, hogy a költségeket alacsonyan tartsa. Ezért a projektmenedzsernek tudnia kell legalább a következőket:

- (1) milyen, a projekthez kapcsolható erőforrásokkal rendelkezik a vállalkozás, (mi az, ami „újrahasznosítható”)?
- (2) hogyan lehet a projektet a lehető legjobban előkészíteni [terminológiai kivonatolás, a terminológiai adatbázis előkészítése, stílusútmutató, a fordítási feladatmeghatározás (*translation brief*) írásba foglalása (Fraser 2000)]?
- (3) hogyan lehet a munkafolyamatot optimálisan megszervezni?

A projekt során a projektmenedzser biztosítja a zökkenőmentes munkavégzést, a dokumentumok szereplők közti mozgatását, és ő válaszolja meg a fordítók és más szereplők kérdéseit is. A projekt lezárultával a projektmenedzser feladata a dokumentum(ok) eljuttatása a fordítás megrendelőjéhez.

A jelenlegi fordítómemóriák azonban nem teszik lehetővé, hogy a projektmenedzser (a továbbiakban: koordinátor) teljes egészében irányítsa a folyamatokat. A koordinátorok ugyan tudják, hogy a projekt szereplőinek pontosan milyen fájlokra van szükségük, nem tudják ezeket betölteni mások alkalmazásaiba. Tapasztalatból és felmérésekből (pl. Drugan 2004) tudjuk azonban, hogy nem minden fordító rendelkezik elég számítástechnikai tapasztalattal ahhoz, hogy a fájl- és beállításkézelés bonyolult feladatával megbirkózzon. Az egygépes fordítómemóriák első hiányossága itt mutatkozik meg.

Ideális munkakörnyezetben a koordinátorok

- (1) erőforrásokat rendelkeznek a felhasználókhoz,
- (2) biztosíthatják, hogy a felhasználók hasznosítsák is az erőforrásokat, az összes erőforrást egy csomagba csomagolhatják,
- (3) beállíthatják a felhasználók fordítómemóriáit.

Milyen erőforrásokat különböztetünk meg? A szereplők segítségével lehetnek a fordítómemóriák, a terminológiai adatbázisok, az eredeti és a munka közben keletkezett dokumentumok, a beállítások (többek között a hálózati beállítások), a validációs szabályok, az XML DTD-k, de a nyelvi erőforrások is. A munkafolyamat-alapú megközelítés révén minden szereplőnek csak annyit kell tennie, amennyi a feladata. A koordinátor biztosítja a projektszintű hatékonyságot, így csak neki kell magas szintű számítástechnikai ismeretekkel rendelkeznie. A fordítók és a lektorok pedig – ha a koordinátor így szeretné – azonnal dolgozni kezhetnek, miután megnyitották az *egyetlen* fájlt, amelyet kaptak. A koordinátor nem állíthatja be mindenki fordítómemóriáját, de biztosíthatja, hogy az összes fordító ugyanazokat az erőforrásokat használja. Az erőforrások frissítése, bővítése esetén a fordítók megkapják az új anyagot. Mindez a kö-

vetkezetesség és a hatékonyság növelése révén megkönnyítheti a technikai tudással nem rendelkező fordító dolgát.

Miután minden személy részére létrehoztunk egy alprojektet, elkezdődik a munkafolyamat. A tapasztalat azt mutatja, hogy a munkafolyamat a legtöbb esetben automatizálható. A munkafolyamatot kezelő motornak pedig – ha azt akarjuk, hogy mindenki használja – minél egyszerűbbnek kell lennie.

A kommunikáció legegyszerűbb és legismertebb módja az e-mail. Korábban kifejlesztettünk egy web- és e-mail-alapú munkafolyamat-automatizálási rendszert, a *Fordítás.net*-et. A rendszerben egy robot intézi az összes levelezést, a fájlokat pedig egy kiszolgálón tárolja, és automatikusan továbbítja a folyamatban soron következő személynek. A felhasználó csak válaszol a levélre, csatolja a munkája eredményét tartalmazó fájlt. A robot minden egyebet elvégez – többek között dokumentálja a teljes folyamatot, így biztosítva a megfelelő minőséget.

A munkafolyamatot átfogó rendszert vertikális hálózati komponensnek tekinthetjük, mivel a különböző szerepeket betöltő személyeket köti össze: a koordinátort a fordítóval, a fordítót a lektorral, a koordinátort az ügyféllel stb. Azonban minél jobban automatizáljuk ezt a rendszert, az árajánlattól az elkészült munka átadásáig és a számlázásig – közben persze bármikor lehetőséget adva a koordinátornak a beavatkozásra –, annál inkább hatékony lesz a rendszer, mert

- (1) a projekt összes erőforrásának központi webes tárhelye megakadályozza a felhasználóknál bekövetkező adatvesztést,
- (2) gyakorlatilag nullára csökken az anyagok továbbításával töltött idő (amikor ember emberrel, a várakozási idő még nagyobb, mint a tranzakcióval töltött idő),
- (3) a fájlnévek részleges és a könyvtárak teljes kiiktatása csökkenti a kihagyás lehetőségét.

A horizontális hálózati komponens az egyforma szerepet betöltő felhasználókat köti össze. Ha a fordítás konzisztenciáját és maximális hatékonyságát biztosítani akarjuk, lehetővé kell tenni, hogy a felhasználók egymással folyamatosan kapcsolatot tarthassanak, és minél hamarabb felhasználhassák a munka közben keletkezett erőforrásokat. A horizontális hálózati komponens magja egy tranzakció-naplózásos erőforrás-frissítő. Az ügyfelek (felhasználói gépek) a koordinátor által megadott időközönként szinkronizálják erőforrásaikat a kiszolgálóval. Ha a koordinátor kicsi értéket ad meg, az eredmény kvázi folyamatos erőforráskövetés.

Látható, hogy ennek a megközelítésnek számos előnye van a kizárólag hálózaton tárolt fordítómemóriával szemben:

- (1) nincs szükség folyamatos hálózati kapcsolatra, de akinek van, élhet annak előnyeivel. Magyarországon becslések szerint csak körülbelül a fordítók fele rendelkezik szélessávú internetkapcsolattal, és a helyi hálózaton kommunikáló belső fordítók is kevesen vannak.
- (2) a hálózati adatforgalom csökken,
- (3) a kiszolgáló számítógép terhelését elosztjuk az ügyfelek között, így egy kiszolgáló több ügyfelet tud kiszolgálni lassulás nélkül.

A kiszolgáló minden ügyfélhez maga kapcsolódik, ezért nem kell beírni a kiszolgáló címét sem, mert a koordinátor állítja be ezt a paramétert. A felhasználók üzeneteket is válthatnak egymással, figyelmeztethetik társaikat a lehetséges problémákra, megjelölhetik a problémás szövegrészeket, terminusokat javasolhatnak és tárgyalhatnak meg (amelyek elfogadás esetén automatikusan be is kerülhetnek a központi adatbázisba), és közvetlenül kommunikálhatnak a koordinátorral.

A horizontális és vertikális hálózati komponensek, az e cikkben nem tárgyalt külső erőforrás-megosztással együtt, optimális teljesítményt biztosítanak hálózati környezetben, és egyszerűsítik a fordítómemóriák használatát.

3. A fájlok és a könyvtárak kiiktatása

A fájlok és a könyvtárak kezelése a fordítók és koordinátorok számára sok gondot okoz. Egy fájl általában az erőforrás egy példányához kapcsolódik, és legfeljebb 255 karakterből álló névvel hivatkozunk rá. A fájlok a fordító számára azonban nem egy, hanem több fontos attribútummal rendelkeznek, és ha az attribútumokat egymás után fűzzük (például egy ilyen elnevezési konvenció alapján: {forrásnyelvkód}{célnyelvkód}{tárgykör-kód}{ügyfélkód}{munkaszám}{egyedi azonosító}) – ez a legjobb, amit tehetünk –, akkor pedig a keresés lesz bonyolult.

A fájlok könyvtárakban találhatók, amelyek vagy egyforma típusú erőforrásokat tartalmaznak, vagy az egy projekthez tartozó erőforrásokat tartalmazzák, a kétféle kategorizáció egyszerre nem lehetséges. Vagy megpróbáljuk megjegyezni, mit hívátettünk, vagy adatbázist készítünk az erőforrásokról, amely minden attribútumot tartalmaz – ehhez jelentős informatikai tudásra van szükség. Esetleg dönthetünk úgy is, hogy megkettőzzük az erőforrásokat, és több tárhelyet használunk fel.

A fájlok és a könyvtárak azért fontosak, mert más szoftverekkel az operációs rendszeren keresztül így lehet érintkezni. Ezért nem lehet tőlük teljesen megszabadulni, de lehetővé tehetjük, hogy a rendszer kezelje őket, és csak egyetlen felületen találkozzunk a fájlrendszerrel.

Az erőforrások természetesen többfélék lehetnek. Minden egyes típushoz meg kell találnunk a leghatékonyabb ábrázolási módot. A fordítómemóriákat és a terminológiai adatbázisokat – amelyeknek természete igen hasonló – lehetséges egyetlen fájlban tárolni. Az ilyen adatbázisok minden egyes bejegyzését el lehet látni többféle attribútummal: ki és mikor készítette a bejegyzést, forrás- és célnyelv, elfogadási státusz, milyen tárgykörbe kerül stb.

A MemoQ-ban a tárgykör (*domain*) kiemelt szerepet kap. A MemoQ a többi fordítómemóriával szemben nem csupán egyetlen tárgykört képes kezelni, hanem az egyes bejegyzéseket többdimenziós térben is el lehet helyezni. Minden tárgykör (pl. autóipar vagy láncatlasz járművek) egy faszerkezet részfája. A különböző tartománydimenziók ortogonálisak, azaz a szövegek különböző szempontok szerinti kategorizálását teszik lehetővé. Annyi tárgykör-dimenziót adhatunk meg, amennyit csak akarunk. Egy nagy fordítóiroda nem csak téma, hanem szervezet (terminológiahasználat stb.), ügyfél (a fordítás megrendelője), és stílus (hivatalos vagy közvetlen stb.) alapján is megadhat tárgyköröket. A tárgykör-információ orientatív, nem pedig restriktív jellegű. Kereséskor azok a szövegrészek jelennek meg először, amelyeknek több tárgyköre egyezik meg a fordítandó szöveg tárgyköreivel. Azok is megjeleníthetők azonban – más szín-

kódolással –, amelyek teljesen más tárgykörbe esnek, de felszíni hasonlóságot mutatnak a fordítandó szöveggel. Így ahelyett, hogy a fordítómemóriákat fájlokban tárolnánk, csak egyetlen fordítómemóriánk és terminológiai adatbázisunk van, amely viszont tárgykörök és más szempontok szerint szűrhető.

A fordítandó dokumentumok hierarchiáját valamilyen módon meg kell őrizni, mivel a projekt befejeztekor a céldokumentumokat exportálni kell az operációs rendszer „világába”. Az információvesztéséget minimalizálendő, az egyazon projekthez tartozó dokumentumokat mindig együtt kell kezelni, és ha a fájlokat szétesztjük, automatikus mechanizmusnak kell visszaállítania az eredeti fájl/könyvtárszerkezetet.

A többi erőforrást, például a DTD-ket, illetve – nyelvi támogatás jelenléte esetén – a szabálybázisokat fájlokként kell kezelni, de a könyvtárstruktúra szükségtelen. A felhasználói felület egyszerűsítése érdekében egyetlen erő-forrásböngészőt kell létrehozni, amelyben a felhasználók (főleg a koordinátorok) láthatják és különböző szempontok alapján szűrhetik az összes erőforrást, és az erőforrásokat projekthez rendelhetik. Így lehet biztosítani, hogy ne veszítsünk el a fordítás során keletkezett értékes erőforrást, és mindent újrahasonlíthassunk.

4. A hatékonyság növelése

A fordítástámogatás e paradigmája szerint minden tevékenységet egyetlen rendszernek kell kezelnie, amelynek fő szervezőeleme a közös munkavégzés. Egyrészt új funkciókra van szükség, másrészt pedig a meglévő funkciókat kell átalakítani.

Az új funkciók beépítésének érdekében testre szabható projekt- és fájlkezelési keretrendszerre és a szabványos erőforrásokat kezelő, lazán kapcsolódó, egymással transzparens módon együttműködő eszközökre van szükség. Ugyanakkor a fordítómemóriák nyelvérzékeny tétele is rendkívüli mértékben megnövelheti a termelékenységet, javíthatja mind a nyelvi eszköz fedését (*recall*), mind pedig pontosságát (*precision*).

Jelenleg a fordítómemóriák *fuzzynak* nevezett algoritmusokkal illesztenek mintákat és tesznek javaslatokat (valójában nem fuzzy logikát alkalmaznak). A morfológiailag egyszerű, többé-kevésbé kötött szórenddel dolgozó nyelvek esetén a statisztika majdnem olyan jól működik, mint az intelligens elemzés, azonban a gazdag morfológiával rendelkező nyelvek esetében egy olyan egyszerű nyelvfüggő művelet, mint a tövesítés is pénzben mérhető jelentős megtakarítást jelenthet. Ha mondatokat is tudunk elemezni, és mondatvázakat létrehozni (Kis-Gröbler-Hodász, 2004), a nyelvtani mintákra is lehet illesztést végezni.

A többdimenziós tartományrendszerrel ellátott intelligens fordítómemória sok kérdést felvet. A felhasználói felületnek egyszerűnek és hatékonynak kell maradnia, mindazonáltal legalább háromféle eredményünk van, amelynek alapján a hasonlóság minőségét értékelni tudjuk. Az első a fuzzy index – a statisztikai hasonlóság –, a második a tartományegyeztési index, amely annál nagyobb, minél több tartomány egyezik meg a forrás- és a hasonló szegmens esetében, a harmadik pedig a nyelvi index, amely nyelvi hasonlóságot keres. A fordítót viszont egyetlen dolog érdekli: a legjobb sorrend. Az indexek összegyűrése egyetlen kombinált indexszé nagy kihívást jelent, és további kutatásoknak ad teret.

Az intelligens elemzés használata a fordítási szegmens fogalmát is értelmetlenné teszi. Ha a számítógép elemezni tudja a mondatot, és kisebb egységeket tud elhatárolni,

természetesen nem a mondat az, ami egy szegmensként számít. Akkor mi? Hogyan lehet a szövegek szinkronizálását (*alignment*) megoldani? Hogyan tudjuk biztosítani a kompatibilitást a többi fordítástámogató eszközzel?

5. Összefoglalás: A MemoQ alapelvei

A MemoQ az itt leírt paradigma első implementációja. Megvalósítjuk benne az összes fenti funkciót, és a szoftvert szorosan integráljuk a már működő munkafolyamat-automatizáló rendszerrel, a Fordítás.net-tel. A MemoQ erősen támogatja a közös munkavégzést a hálózati komponensek és az egyszerű, fájlműveletek nélkül működő fordítói felhasználói felület segítségével. Az első nyelvileg támogatott nyelvpár az angol-magyar lesz, amely a MorphoLogic fordítómemória-kutatására és a MorphoLogic kutatói által kifejlesztett MorphoTM fordítómemória-motorra épül. A MemoQ azonban a jelenlegi fordítómemóriákhoz hasonlóan nyelvfüggetlenül is képes működni, hagyományos hasonlósági keresés révén. A MemoQ első verziója 2005 őszére várható.

Elmondhatjuk, hogy nincs átmenet az egygépes alkalmazások és a közös munkavégzést támogató alkalmazások között. A meglévő programokat nem lehet átírni úgy, hogy teljesen megfeleljenek a fenti elvárásoknak. Az új paradigma alapján új fordítómemóriákat kell írni, vagy a régieket teljesen újraírni. Az új megközelítésnek azonban a régi részhalmaza, ezért lehetséges a problémamentes fokozatos áttérés az egygépes megközelítésről a közös munkavégzésre.

6. Köszönetnyilvánítás

A szerzők köszönetet mondanak a MorphoLogic kutatóinak a MorphoTM nyelv-érzékeny fordítómemória-technológia rendelkezésre bocsátásáért, illetve az ActiWise Kft.-nek – különösen pedig Cserép Csabának – a Fordítás.net rendszer kifejlesztésében nyújtott segítségéért.

Irodalom

- DRUGAN, J. (2004): Training Tomorrow's Translators, Universidad Europea de Madrid. In *Proceedings of Jornadas de Traducción e Interpretación*, Madrid. (megjelenés előtt)
- FRASER, J. (2000): The broader view: how freelance translators define "translation competence", In: Schäffner and Adab (eds.), *Developing Translation Competence*, John Benjamins, Amsterdam, pp. 51-62
- FULFORD, H., GRANELL-ZAFRA, J. (2004): The freelance translator's workstation: an empirical investigation, In: *Proceedings of the Ninth EAMT Workshop*, University of Malta, Valletta, pp. 53-61.
- HODÁSZ, G., GRÖBLER, T., KIS, B.: Translation memory as a robust example-based translation system, In: *Proceedings of the Ninth EAMT Workshop*, University of Malta, Valletta (2004) pp. 82-89.
- KIS Balázs, LENGYEL István (2003): Új módszerek az emberi fordítás gépi támogatásában. In: *Az I. Magyar Számítógépes Nyelvészeti Konferencia gyűjteményes kötete*, Szegedi Tudományegyetem, Szeged.

- REINEKE, D. (2004): Localización y terminología: efectos sinérgicos obtenidos a raíz de un experimento realizado en el aula. Universidad Europea de Madrid, Madrid (*megjelenés előtt*)
- RICO PÉREZ, C. (2002): Translation and Project Management. In: *Translation Journal*, 6, 4. <http://accurapid.com/journal/22project.htm>
- ROBICHAUD, B., L'HOMME, M-C. (2003): Teaching the automation of the translation process to future translators, <http://www.dlsi.ua.es/t4/docum/robichaud.pdf>